Supplementary Material

1. Integral Feature set conversion

Table S.1 Comprehensive analysis of each classifier for MDVR-KCL(Read Text).

ML Model	Feature Set	80:20 ratio									5-fold (%)	10-fold (%)
		Accuracy (%)	Precision	Recall	F1 Score	ROC AUC	MCC	Cohen's Kappa	Training Time	Testing Time		
RF	MFCC	81.80%	0.852	0.650	0.738	0.94	0.62	0.60	156.546	0.038	85.17	84.77
	Spectrogram	77.45%	0.730	0.675	0.701	0.84	0.52	0.52	94.390	0.122	76.52	76.72
	Spectral	81.37%	0.828	0.663	0.736	0.89	0.60	0.59	132.747	0.040	74.36	74.75
	Voice_Quality	100.00%	1.000	1.000	1.000	1.00	1.00	1.00	66.495	0.072	99.90	99.02
	All Combined	98.53%	0.975	0.988	0.981	1.00	0.97	0.97	144.121	0.116	98.33	98.13
KNN	MFCC	89.71%	0.915	0.813	0.861	0.96	0.78	0.78	2.974	0.007	90.47	90.96
	Spectrogram	65.20%	0.605	0.325	0.423	0.68	0.22	0.20	0.592	0.010	63.36	63.65
	Spectral	63.24%	0.533	0.500	0.516	0.66	0.22	0.22	0.467	0.004	64.64	62.97
	Voice_Quality	98.04%	0.987	0.963	0.975	1.00	0.96	0.96	0.439	0.003	91.35	95.09
	All Combined	74.50%	0.700	0.613	0.653	0.72	0.46	0.45	4.233	0.017	66.71	67.97
DT	MFCC	70.10%	0.604	0.688	0.643	0.75	0.39	0.39	8.902	0.014	70.34	73.27
	Spectrogram	73.53%	0.732	0.513	0.603	0.76	0.43	0.41	1.913	0.004	70.92	68.66
	Spectral	71.57%	0.641	0.625	0.633	0.74	0.40	0.40	2.674	0.004	72.30	71.42
	Voice_Quality	100.00%	1.000	1.000	1.000	1.00	1.00	1.00	1.324	0.004	94.70	94.41
	All Combined	100.00%	1.000	1.000	1.000	1.00	1.00	1.00	5.455	0.002	82.80	86.64
Adaboost	MFCC	83.82%	0.831	0.738	0.781	0.92	0.66	0.65	332.499	0.063	79.76	78.58
	Spectrogram	78.43%	0.737	0.700	0.718	0.82	0.54	0.54	57.863	0.069	68.47	66.30
	Spectral	81.37%	0.828	0.663	0.736	0.90	0.60	0.59	89.127	0.076	69.94	69.15
	Voice_Quality	100.00%	1.000	1.000	1.000	1.00	1.00	1.00	9.564	0.002	97.35	97.05
	All Combined	100.00%	1.000	1.000	1.000	1.00	1.00	1.00	183.142	0.040	98.43	96.66
Ensemble model	MFCC	**90.59%**	**0.88**	**0.85**	**0.87**	**1.00**	**0.78**	**0.78**	**115.64**	**0.02**	80.16	80.84
	Spectrogram	78.94%	0.72	0.71	0.72	1.00	0.54	0.54	53.91	0.02	75.83	75.84
	Spectral	84.82%	0.84	0.73	0.78	1.00	0.66	0.65	69.20	0.04	75.05	76.63
	Voice_Quality	**100.00%**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**34.86**	**0.02**	100.00	100.00
	All Combined	98.03%	0.96	0.99	0.98	1.00	0.96	0.96	211.58	0.05	96.36	96.47

*Note: Bold value indicates the best performance

Table S.2 Comprehensive analysis of each classifier for PC-GITA (Read Text).

ML Model	Feature Set	Accuracy (%)	Precision	Recall	F1 Score	ROC AUC	MCC	Cohen's Kappa	Training Time	Testing Time	5-fold (%)	10-fold (%)
RF	MFCC	44.444	0.421	0.364	0.390	0.491	-0.116	-0.115	214.609	0.188	56.121	56.542
	Spectrogram	62.222	0.609	0.636	0.622	0.666	0.245	0.245	79.742	0.307	50.283	50.198
	Spectral	55.556	0.556	0.455	0.500	0.580	0.109	0.107	84.275	0.133	63.293	63.715
	Voice_Quality	80.000	0.882	0.682	0.769	0.931	0.613	0.598	96.893	0.262	74.212	76.996
	All Combined	66.667	0.667	0.636	0.651	0.779	0.333	0.332	147.011	0.084	60.596	62.055
KNN	MFCC	57.778	0.571	0.545	0.558	0.636	0.154	0.154	6.275	0.100	57.899	55.672
	Spectrogram	51.111	0.500	0.500	0.500	0.573	0.022	0.022	4.842	0.009	45.222	49.723
	Spectral	35.556	0.316	0.273	0.293	0.385	-0.296	-0.293	4.003	0.015	42.960	48.399
	Voice_Quality	84.444	0.857	0.818	0.837	0.864	0.689	0.688	4.629	0.010	61.081	61.996
	All Combined	48.889	0.471	0.364	0.410	0.496	-0.029	-0.028	5.164	0.033	45.677	46.581
DT	MFCC	56.790	0.576	0.475	0.521	0.639	0.136	0.134	11.093	0.005	53.846	56.567
	Spectrogram	60.494	0.591	0.650	0.619	0.658	0.212	0.211	0.991	0.000	49.358	54.317
	Spectral	61.728	0.655	0.475	0.551	0.655	0.241	0.232	1.124	0.005	58.577	52.323
	Voice_Quality	95.062	0.929	0.975	0.951	0.985	0.902	0.901	0.985	0.004	84.102	86.329
	All Combined	77.778	0.762	0.800	0.780	0.784	0.556	0.556	2.127	0.021	65.506	64.793
Adaboost	MFCC	56.790	0.564	0.550	0.557	0.583	0.135	0.135	137.875	0.304	56.824	58.543
	Spectrogram	50.617	0.500	0.475	0.487	0.528	0.012	0.012	157.958	0.263	53.086	49.098
	Spectral	61.728	0.600	0.675	0.635	0.592	0.237	0.236	141.733	0.234	57.806	58.098
	Voice_Quality	98.765	0.976	1.000	0.988	0.998	0.976	0.975	142.180	1.607	78.127	92.256
	All Combined	76.543	0.733	0.825	0.776	0.825	0.536	0.532	388.858	0.787	66.478	67.494
Ensemble model	MFCC	56.790	0.568	0.525	0.545	0.564	0.135	0.135	21.889	0.189	59.287	56.579
	Spectrogram	55.556	0.577	0.375	0.455	0.569	0.114	0.107	2.824	0.112	52.361	45.683
	Spectral	64.198	0.720	0.450	0.554	0.707	0.302	0.281	3.152	0.094	53.346	51.878
	Voice_Quality	97.531	0.975	0.975	0.975	0.995	0.951	0.951	2.462	0.063	95.540	97.524
	All Combined	88.889	0.844	0.950	0.894	0.927	0.784	0.778	4.265	0.141	63.065	79.366

2. Spectral feature Set:

Table S.3. Comprehensive analysis of Performance evaluation for MDVR-KCL dataset.

10-fold cross validation(mean)											5-fold cross validation
Models	Feature Set	Accuracy (%)	Precision	Recall	F1 Score	ROC AUC	Cohen's Kappa	Specificity	Training Time (s)	Testing Time (s)	Accuracy (%)
MobileNetV2	CQT	94.09	0.94	0.94	0.94	0.97	0.87	0.96	157.34	4.66	69.90
	Chroma	88.29	0.88	0.88	0.88	0.94	0.74	0.94	149.25	4.16	66.06
	MFCC	80.95	0.81	0.81	0.80	0.87	0.57	0.92	145.56	4.74	76.85
	Mel	92.67	0.93	0.93	0.92	0.95	0.84	0.97	208.67	12.51	89.04
	STFT	88.00	0.88	0.88	0.88	0.94	0.74	0.94	208.67	12.51	65.80
VGG-19	CQT	81.62	0.79	0.82	0.80	0.86	0.57	0.93	3658.81	12.98	90.09
	Chroma	77.33	0.77	0.77	0.77	0.83	0.50	0.87	1635.71	7.49	71.05
	MFCC	70.76	0.71	0.71	0.68	0.76	0.31	0.89	2554.37	12.32	67.61
	Mel	81.05	0.81	0.81	0.81	0.87	0.58	0.89	1758.83	13.34	78.95
	STFT	75.24	0.77	0.75	0.72	0.81	0.41	0.92	6775.18	6.96	73.90
ResNet50	CQT	74.29	0.74	0.74	0.72	0.76	0.40	0.91	8268.41	8.03	75.03
	Chroma	69.81	0.69	0.70	0.67	0.74	0.29	0.88	1135.74	8.31	61.05
	MFCC	70.76	0.71	0.71	0.68	0.76	0.31	0.90	1048.36	7.18	62.57
	Mel	62.57	0.39	0.63	0.48	0.59	57.00	1.00	4397.08	7.18	62.57
	STFT	70.29	0.66	0.70	0.63	0.70	0.25	0.96	1048.36	7.18	89.90
Proposed Model (Dataset-I)	CQT	**93.90**	**0.939**	**0.939**	**0.937**	**0.960**	**0.864**	**0.969**	**12325.36**	**7.04**	**88.57**
	Chroma	**92.10**	**0.924**	**0.921**	**0.918**	**0.946**	**0.823**	**0.965**	**6305.56**	**8.17**	**85.8**
	MFCC	**81.90**	**0.827**	**0.819**	**0.805**	**0.877**	**0.585**	**0.921**	**1907.31**	**5.78**	**73.04**
	Mel	**94.19**	**0.942**	**0.942**	**0.941**	**0.966**	**0.872**	**0.971**	**2756.17**	**4.50**	**89.04**
	STFT	**95.90**	**0.961**	**0.959**	**0.958**	**0.978**	**0.908**	**0.988**	**2718.81**	**4.87**	**88.47**

Table S.4. Comprehensive analysis of Performance evaluation for Image spectrum for PC-GITA (Read Text).

10-fold cross validation(mean)											5-fold
Models	Feature Set	Accuracy (%)	Precision	Recall	F1 Score	ROC AUC	Cohen's Kappa	Specificity	Training Time (s)	Testing Time (s)	Accuracy (%)
MobileNetV2	CQT	95.20	0.95	0.95	0.95	0.97	0.90	0.96	157.34	4.66	71.42
	Chroma	66.55	0.69	0.67	0.66	0.74	0.33	0.76	149.25	4.16	84.08
	MFCC	85.35	0.86	0.85	0.85	0.93	0.71	0.84	145.56	4.74	69.38
	Mel	96.40	0.97	0.96	0.96	0.97	0.93	0.95	684.47	4.20	79.18
	STFT	97.20	0.97	0.97	0.97	0.97	0.94	0.98	208.67	12.51	75.51
VGG-19	CQT	80.29	0.78	0.80	0.78	0.83	0.54	0.93	3658.81	12.98	90.09
	Chroma	78.67	0.78	0.79	0.78	0.85	0.53	0.88	1635.71	7.49	75.33
	MFCC	71.14	0.71	0.71	0.68	0.78	0.32	0.89	2554.37	12.32	69.14
	Mel	81.52	0.81	0.82	0.81	0.86	0.62	0.89	1758.83	13.34	51.41
	STFT	88.59	0.88	0.89	0.88	0.89	0.56	0.91	6775.18	6.96	66.12
ResNet50	CQT	62.22	0.63	0.62	0.61	0.64	0.23	0.47	8268.41	8.03	53.87
	Chroma	66.55	0.69	0.67	0.66	0.74	0.33	0.76	1135.74	8.31	52.65
	MFCC	52.22	0.27	0.52	0.36	0.60	0.20	0.45	476.45	7.18	51.42
	Mel	56.30	0.56	0.56	0.54	0.61	0.11	0.37	4397.08	7.18	55.51
	STFT	53.93	0.50	0.54	0.48	0.56	0.06	0.67	1048.36	7.18	51.42
Proposed Model	CQT	**97.20**	**0.97**	**0.97**	**0.97**	**0.99**	**0.94**	**0.98**	**403.32**	**1.21**	**88.57**
	Chroma	91.57	0.92	0.92	0.92	0.95	0.83	0.88	608.95	1.24	85.8
	MFCC	86.30	0.87	0.86	0.86	0.90	0.72	0.80	725.61	3.31	73.04
	Mel	92.59	0.93	0.93	0.92	0.93	0.85	0.89	503.87	1.26	89.04
	STFT	93.20	0.94	0.93	0.93	0.97	0.87	0.91	7102.88	1.22	88.47

*Note: Bold value indicates the best performance

3. Independent Speaker Validation: Dataset-I, II

Table S.5: Performance evaluation metrics of independent speakers for dataset-I using different Feature Set.

Feature Set	Accuracy (%)	Precision	Recall	F1 Score	Training Time (s)	Testing Time (s)
CQT	79 ± 6	0.80 ± 0.05	0.79 ± 0.06	0.78 ± 0.07	5867	32.9
Chroma	73 ± 5	0.75 ± 0.04	0.73 ± 0.05	0.71 ± 0.08	3912	9.0
MFCC	77 ± 8	0.78 ± 0.08	0.77 ± 0.08	0.76 ± 0.08	4293	14.8
Mel	75 ± 5	0.77 ± 0.03	0.75 ± 0.05	0.73 ± 0.07	7239	19.8
STFT	63 ± 4	0.53 ± 0.13	0.63 ± 0.04	0.56 ± 0.08	4674	19.3

Table S.6: Performance evaluation metrics of independent speakers for dataset-II using different Feature Set.

Feature Type	Accuracy (%)	Precision	Recall	F1-Score	Training Time (s)	Testing Time (s)
CQT	68 ± 5	0.63 ± 0.14	0.68 ± 0.05	0.63 ± 0.10	5086	8.9
Chroma	61 ± 2	0.37 ± 0.02	0.61 ± 0.02	0.46 ± 0.02	1358	7.5
MFCC	63 ± 4	0.53 ± 0.13	0.63 ± 0.04	0.56 ± 0.08	4674	19.3
Mel	67 ± 5	0.68 ± 0.07	0.67 ± 0.05	0.65 ± 0.07	8262	29.3
STFT	77 ± 5	0.80 ± 0.03	0.77 ± 0.05	0.75 ± 0.08	3680	8.5